Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 45 (2012) 815 – 823

Procedia Engineering

www.elsevier.com/locate/procedia

2012 International Symposium on Safety Science and Technology

Identification of Hazardous Road Locations for Pedestrians

Shenjun YAO, Becky P. Y. LOO *

Department of Geography, The University of Hong Kong, Pokfulam, Hong Kong

Abstract

Hot zone methodology is promising in the identification of hazardous road locations. The key steps involve geo-validation of road crashes, segmentation of road network into basic spatial units (BSUs), calculation of actual crash intensity, definition of threshold value, and examination of spatial proximity of BSUs. This research applies the hot zone methodology to identify dangerous road locations for pedestrians during the study period of 2005 to 2007 in Kwun Tong District of Hong Kong. In particular, the crash intensity was calculated by a casualty-weighted method, which assigns different weights to different injury severity types. Two negative binomial regression models were employed to determine the threshold values. One could be treated as a base model which includes the length of BSU as the only explanatory variable. The other, regarded as a full model, introduces diverse environmental variables that might have influenced the distribution of pedestrian casualties.

Keywords: injury; road crash; hot zone; pedestrians; safety

1. Introduction

Pedestrians are regarded as the most vulnerable road user type. In many low-income and middle-income countries, the share of pedestrian fatalities in total road deaths is alarmingly high. Even in high-income and motorization countries where people may drive more than walk, the percentage of pedestrian fatalities is still relatively high. According to the European Commission (2010), 7,638 pedestrians died in road traffic crashes in 23 European countries in 2008, accounting for over 20% of road traffic fatalities in these countries. In Hong Kong, pedestrians are at high crash risks. Although the shares of pedestrians in total injuries were decreasing slightly from 25% to 20% during the period from 2001 to 2010, pedestrians were still the most vulnerable road user group in terms of fatal and serious injuries. In 2010, there were 117 fatalities and 2,160 serious injuries, of which 59% and 34.9% were pedestrians. The rate of fatal and serious injuries in pedestrian casualties was almost three-fold that in driver and passenger casualties. Such alarming situation merits both academic and public's concerns.

The first step to improve pedestrian safety can be the identification of hazardous road locations for pedestrians. Hot zone methodology is a promising method in identifying dangerous road locations. Unlike traffic hotspot (also known as blacksite or blackspot) identification which takes junctions or individual road segments as spatially independent units [1-3], the hot zone methodology takes network continuity into consideration for identifying concentrations of road crashes [4-7]. The key steps of hot zone methodology involve the geo-validation of road crashes, segmentation of road network into basic spatial units (BSUs), calculation of actual crash intensity, definition of threshold value and spatial modeling of crash patterns [8-9]. Previous research on hot zone identification only concentrated on crash frequency. However, the analysis on casualties is also important as medical, social and other costs of traffic crashes are arguably much more closely related to the number of

* Corresponding author. Tel.: +0-852-391-77024; fax: +0-852-255-98994.
E-mail address: bpyloo@hku.hk

doi:10.1016/j.proeng.2012.08.244

casualties, rather than the number of crashes. For instance, fatal and severely injured victims require the timely dispatch of ambulances and medical treatments by trauma teams in hospitals. Investigation of the spatial pattern of pedestrian casualties can help administrations to allocate the resources more appropriately. Hence, this study targets traffic injury casualties and analyzes the spatial distribution of pedestrian victims. In addition, previous studies on hot zone identification mainly used numerical definitions such as an arbitrary number [6] or statistical definitions like Monte Carlo Simulation with certain significance levels [7] to determine the threshold value. This study makes attempts to employ a model-based definition which introduces crash prediction models to estimate the expected number of pedestrian casualties as the threshold value for hot zone identification.

2. Hot zone methodology

This section briefly introduces the general steps of hot zone identification. Details of the procedure can also be found in Loo and Yao [8].

2.1. Geo-validation of road crashes

As a 1D phenomenon, traffic crashes are constrained to the road network. However, for both technical and non-technical reasons, a great number of reported road crashes cannot intersect with the centerlines of the road network [10]. As high precision of locations of road crashes is vital to the identification of hazardous road locations, the road crashes were first geo-validated and located on the centerlines of road network before the spatial analysis [10].

2.2. Segmentation of road network into BSUs

Although there has been no clear indication of an optimal length for defining a hazardous BSU, many researchers have suggested using a fixed value which should be short enough to allow the identification of crash clusters but long enough to reflect variations in road environment [6]. In this paper, the length of a BSU is set to be 100 meters. However, for an empirical link-node road system, the length of each BSU still varies substantially after the standard segmentation procedures. The main reason is that the "fixed length" condition is likely to be violated near end nodes and resulted in some BSUs having less than the predefined length. In order to reduce the number of short BSUs, Loo and Yao [8] developed a dissolving algorithm to transform a raw-link-node-road system (Fig. 1a) to a dissolved-road one (Fig. 1b) before segmentation. They applied the algorithm to the entire road network of Hong Kong. The results suggested that the dissolving algorithm could result in a sharp decline of short BSUs.

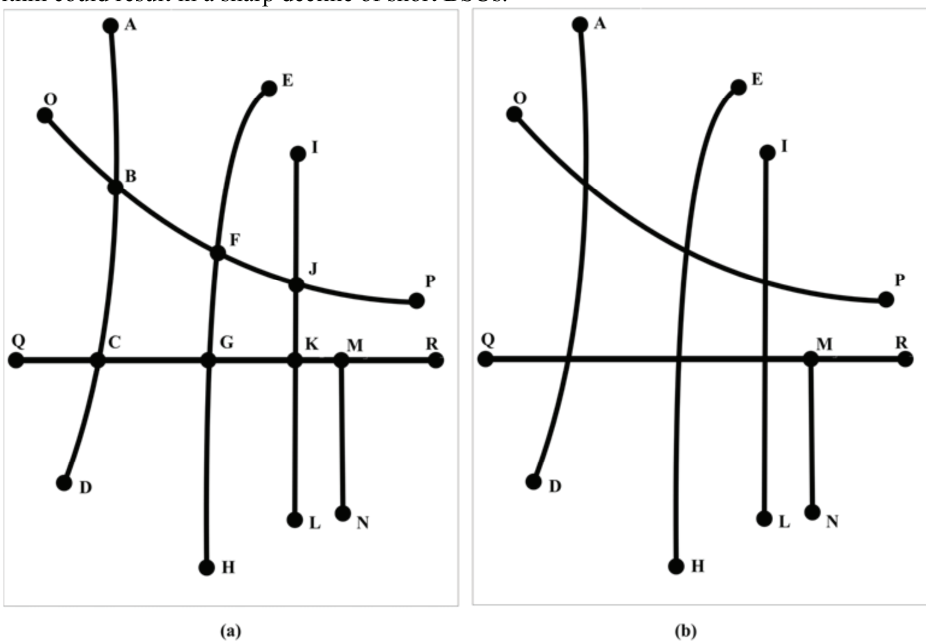


Fig. 1. Illustration of road structures for (a) a raw link-node system and (b) a dissolved road system.

2.3. Calculation of actual crash intensity

As mentioned earlier, most previous research calculated the actual crash intensity (ACI) by means of crash frequency, which aggregates road crashes by BSU and treats the number of observed road crashes as the value of ACI. This research introduces a casualty-based cost-weighting approach to assign different weights to different injury severity levels [11].

2.4. Definition of threshold values

Although there is no consensus on the definition of threshold values for hot zone identification, there are three common types of definitions for hot spot identification, that is, numerical, statistical and model-based [1][12]. The simple numerical definitions based on the recorded number of traffic crashes were still the mostly commonly adopted by administrations worldwide [6][11][13]. The second group uses a statistical definition based on a "normal" or "expected" value of a similar type of targeted locations. A road segment having an actual crash intensity measure (statistically) significantly higher than the "expected" value is considered dangerous. Finally, model-based definitions use thresholds from crash prediction models which take into account more statistically confounding factors.

2.5. Modeling of the spatial pattern of BSUs

An indicator $I_{(HZ)}$ is introduced by Loo and Yao [8] to model the spatial crash pattern, which can be calculated by:

$$I_{(HZ)i} = z_i \sum_{j=1, j\neq i}^{n} W_{ij} z_j \tag{1}$$

where n is the number of BSUs; i, j=1,2,…,n, W_{ij} is a contiguity 0-1 matrix; Z_i is a 0-1 indicator showing whether or not the BSU is "hot". Here, "hot" means the BSU has an ACI greater than the threshold value. Z_i can be denoted by:

$$z_i = \begin{cases} 1 & if\ a_i > t_i \\ 0 & otherwise; \end{cases}, \tag{1}$$

where a_i is the ACI at BSU i; t_i is the threshold value of BSU_i ,which can be assigned any positive value.

The value of $I_{(HZ)i}$ can be zero or a positive integral number between 1 and n-1. A zero value may indicate that the ACI is no more than the threshold value at BSU i, or that the BSU i itself has an ACI higher than the threshold value but there is no other contiguous hot BSU in the vicinity. And a positive value means that both the BSU i and at least one of its contiguous BSUs have their ACIs greater than the threshold value.

3. Identification of hazardous road locations in Kwun Tong District

3.1. Study area

Kwun Tong District of Hong Kong is selected as the study area for the identification of hazardous road locations of pedestrians because it had the second highest absolute number of population (58,742) and ranked first in terms of population density (0.52 persons per square meters) among 18 districts of the city (see Fig. 2) based on the 2006 Census data (Census and Statistics Department 2007).

3.2. Geo-validation of pedestrian casualties

The police crash investigation data on traffic crashes, known as the Traffic Road Accident Database (TRADS), provides us a vast range of information on each crash. This system records the location of crashes, road user type and the injury severity classification of casualties, which enables us to locate casualties, identify pedestrian victims and classify injury into fatal, slight and serious types. Based upon the geographical coordinates, casualties were firstly plotted onto a map and were geo-validated by following a geo-validation process introduced by Loo [10]. As road crashes are rare events and randomness in the number of casualties happening at a certain BSU is typical, it is of great importance that the study period can ensure representative casualty samples [7]. Hence, following Mueller, Rivara, and Bergman [14], this study pooled the datasets from 2005 to 2007 into a three-year period for identifying traffic hot zones. There were altogether 940 pedestrian casualties during the period, of which 21 were fatal and 242 were severely injured.

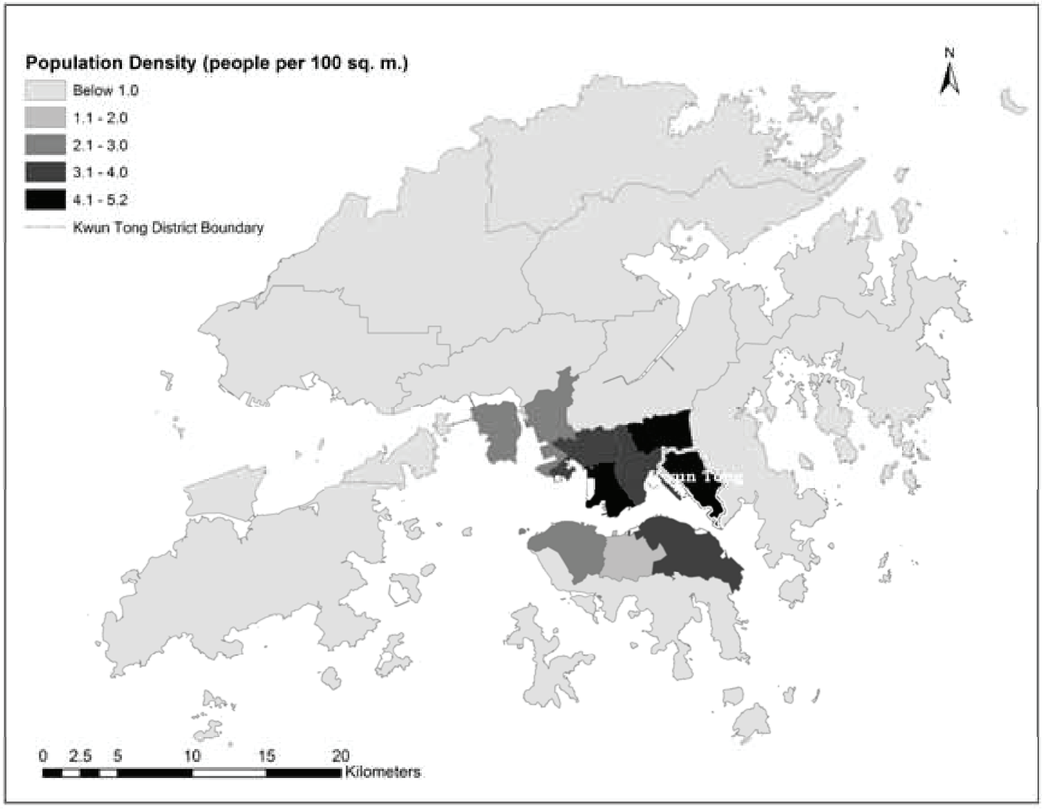


Fig. 2. Study area.

3.3. Segmentation of road network

The road network in Kwun Tong District is 164 km long with 1,357 links. After the dissolving procedures (see Fig. 1), the number of links was reduced to 519. The dissolved road system was then segmented into BSUs with 100 m interval. Finally, there were altogether 1,852 BSUs, of which about 25% were less than 100 m long.

3.4. Calculation of costs

The cost-weighting method is used for computing ACI, which assigns different weights to different severity levels based on their costs. The key step of this procedure is to determine the costs of a road crash fatality, a serious injury and a slight injury. Research into the benefit of preventing a road crash fatality, which is often measured by the Value of a Statistical Life (VSL), has a long history. Several techniques have been employed to estimate VSL, including the Human Capital (HC), Willingness-to-pay (WTP), Stated Preference (SP) and Revealed Preference (RP) approaches. Comprehensive reviews of these methods exist in the economic and road safety literature [15-16]. With advances in the methodology of WTP, a growing number of countries have preferred WTP in recent years. The approach estimates the value of preventing a fatality by estimating the amount of money that individuals would be prepared to pay to reduce the risk of loss of life [17]. While some studies have been conducted to establish WTP values for traffic fatalities, those that also estimated values for serious or slight injuries were limited. This is mainly because of problems of designing reliable and valid questionnaires to elicit WTP estimates [17]. Nevertheless, there are some developed countries employing the WTP method to estimate the costs of non-fatal casualties. Table 1 lists some of these WTP casualty costs. On average, the costs of a serious injury and a slight injury are approximately 16% and 1% of the value of a fatality respectively. With the absence of casualty cost data in Hong Kong, this research is based on these average ratios and defines the cost-weighted crash intensity as:

$$\lambda_{(i)} = 100Fa_{(i)} + 16Se_{(i)} + Sl_{(i)} \tag{3}$$

where i=1,2,3…n; n is the number of BSUs; Fa is the number of fatalities; Se is the number of serious injuries; and Sl is the number of slight injuries.

Table 1. Casualty cost in developed countries using willingness-to-pay approach

	Year	Currency	Cost			Ratio	
			Fatal	Serious	Slight	Serious/Fatal	Slight/Fatal
Austria	2006	€	2,676,374	316,772	22,722	11.9%	0.8%
New Zealand	2006	NZ$	3,065,000	535,000	60,000	17.5%	2.0%
Singapore	2008	S$	1,874,000	243,600	18,740	13.0%	1%
Sweden	2005	SK	18,383,000	3,280,000	N/A	17.8%	N/A
United Kingdom	2006	£	1,489,163	167,332	12,898	11.2%	0.9%
United States	2007	$	5,800,000	1,322,300	30,920	22.7%	0.5%
Average						15.6%	1.0%

3.5. Definition of threshold values

The negative binomial model was employed to determine the threshold value. The expected number of fatalities or serious injuries or slight injuries is estimated by means of negative binominal regression, λ_i which is denoted by:

$$\ln(\lambda_i) = \beta' x_i + \varepsilon \tag{4}$$

where $\ln(\lambda_i)$ is the expected natural log of casualties, x_i is a vector of predictors and β' are the estimated coefficients.

Two negative binomial regression models were used to determine the threshold values. One could be treated as a base model which includes the length of BSU as the only explanatory variable. The other, regarded as a full model, introduces environmental variables that may affect the risk of pedestrian casualties.

- Base Model

Three sub-models were established for fatalities, serious injuries and slight injuries respectively. Nonetheless, the independent variable is length of BSU only.

- Full Model

The concentration of road crashes is not driven by a single force, but a combination of contributory factors. Among different factors, environmental indicators, such as land use, socio-economic, demographic and road structure characteristics, have been closely examined by researchers [18-21]. The explanatory variables used in this research were chosen based upon previous studies. We included some variables at the Tertiary planning unit (TPU) level as characteristics of the neighborhoods may affect the chance of being involved in a traffic crash. In Hong Kong, TPUs are devised by the Hong Kong Planning Department for town planning purposes and are used for population census. The whole territory of 1,104 square kilometers in Hong Kong was divided into 282 TPUs in 2006, of which 11 were located in Kwun Tong District. For each TPU-based variable, all the BSUs within a same TPU were assigned the same value.

Similar to the base model, three sub-models were established for three different injury types. For each sub-model, in addition to length of BSU, the independent variables also include:

<u>Number of road junctions on BSU (link level)</u>

The road junctions are aggregated by BSU.

<u>Land use mix (TPU level)</u>

Land use mix is calculated based on the Simpson Diversity Index. It is a biological diversity measurement which evaluates the number of land use categories within a neighborhood. Following the formal expression of the index, the land use diversity, denoted as LUD, can be calculated by:

$$LUD = \frac{\sum n_i(n_i - 1)}{N(N-1)} \tag{5}$$

where N is the total number of land use and n_i is the number of land use in the ith category. The land use data were extracted from a paper map on land utilization of Hong Kong, which was compiled from the 2006 land use data of the Planning Department and other relevant information including data derived from SPOT Satellite images (Planning Department, 2007). The transformation from a paper map to a digitized vector map was performed by using ENVI which is the premier software for processing and analyzing geospatial imagery. The key steps are to:

(a) digitize the paper map;

(b) select samples for each type of land use;

(c) perform Supervised Classification;
(d) merge smaller polygons;
(e) vectorize the raster data to shape files (ArcGIS Vector Format); and
(f) manually check and correct the classification errors.

The land use data is then broadly divided into five categories (residential, commercial, industrial, institutional and others).

Socio-economic deprivation index (TPU level)

The socio-economic deprivation index was derived from the 2006 By-census. The Census and Statistics Department of Hong Kong conducts a population census every ten years and a by-census in the middle of the intercensal period. In this study, some indicators were extracted to describe socio-economic conditions of each TPU from the 2006 By-census reports. The variables are based on previous literature on area deprivation predictors of pedestrian casualties. The indicators on income, owner-occupancy, education, occupation and unemployment are defined as follows:

(a) income: monthly household income <6000 HKD (%);
(b) owner-occupancy: not owner-occupied households (%);
(c) education: low upper-secondary education attainment (%);
(d) occupation: occupation with no or low qualifications (%); and
(e) unemployment: unemployment (%).

Depending on these five predictors, the socio-economic deprivation index (SDI) was calculated using the Z-scores method, which is given as:

$$SDI_i = \sum_{j=1}^{m} W_j Z_{ij} \tag{6}$$

$$Z_{ij} = \frac{V_{ij} - \mu_j}{\sigma_j} \tag{7}$$

where SDI_i is the socio-economic deprivation score for the ith TPU, i=1,2,3,…,n, n is the count of TPUs, m is the number of variables, V_{ij} is the actual value of jth variable for TPU i, μ_j is the mean value of variable j, σ_j is the standard deviation of jth indictor, and W_j is the weight attached to z-scores. The weights were produced using the first principal-component analysis. The details of the calculation can be found in Loo and Yao [21].

4. Results

The descriptive statistics on dependent and independent variables are shown in Table 2; and the results on negative binomial regression models are summarized in Table 3. It can be observed that the length of BSU and number of junctions on BSU were significantly related with the dependent variables with p-values less than 0.05, regardless of the models used. Land mix and SDI were more associated with the occurrence of serious and slight injuries. Using these models, the expected numbers of fatal, severely and slightly injured pedestrians were estimated. Following Equation 3, the expected costs were then calculated as the threshold values for identifying the pedestrian hot zones.

Table 2. Descriptive statistics on dependent and independent variables

Type	Mean	Median	Mode	Std. Deviation	Minimum	Maximum
Dependent variables						
Number of pedestrians killed	0.01	0.00	0.00	0.12	0	2
Number of serious injured pedestrians	0.13	0.00	0.00	0.45	0	6
Number of slight injured pedestrians	0.37	0.00	0.00	0.92	0	8
Independent variables						
Length of the BSU	84.84	100.00	100.00	27.46	1.06	100.55
Number of junctions on the BSU	0.99	1.00	1.00	0.83	0	6
Land mix	25.36	22.60	36.70	19.87	0.30	52.00
SDI	1.57	1.69	-0.36	2.97	-12.93	17.48

Table 3. Results on negative binomial regression models

Type	Intercept	Length of BSU	Number of junctions on BSU	Land Mix	SDI
Pedestrians killed					
Base model	-7.839**	0.36***	-	-	-
Full model	-9.189***	.047**	.523**	.005*	.240*
Seriously injured pedestrians					
Base model	-4.858***	0.31***	-	-	-
Full model	-7.458***	.038***	.359***	.013*	.155**
Slightly injured pedestrians					
Base model	-3.692***	0.29***	-	-	-
Full model	-5.105***	.034***	.236***	.027***	.347***
***: p-value <0.01; **: p-value <0.05; *: p-value <0. 1					

Statistics on the pedestrian hot zones are shown in Table 4. The results demonstrate that the full model detected more and longer hot zones than the base model, whereas the base-model hot zones had larger number of casualties with more costs. Figure 3 depicts the locations of hot zones, with the top five pedestrian hazardous locations (based on total costs) highlighted. If the two types of pedestrian hot zones are overlaid, the hot zones detected by both models and by one model only can be identified. There were 40 pedestrian hot zones identified by both models. These hot zones should be treated as top priority. Five out of fifty were only identified in the base model and ten were only detected in the full model. These hot zones are also worth investigations. For instance, hospitals may be more interested in hot zones with threshold values determined by the base model, rather than the full model, as they are more concerned with actual number/cost of injuries. However, traffic engineers may like to know more about hot zones with threshold values determined by filtering some “prior information” for treatment and evaluation.

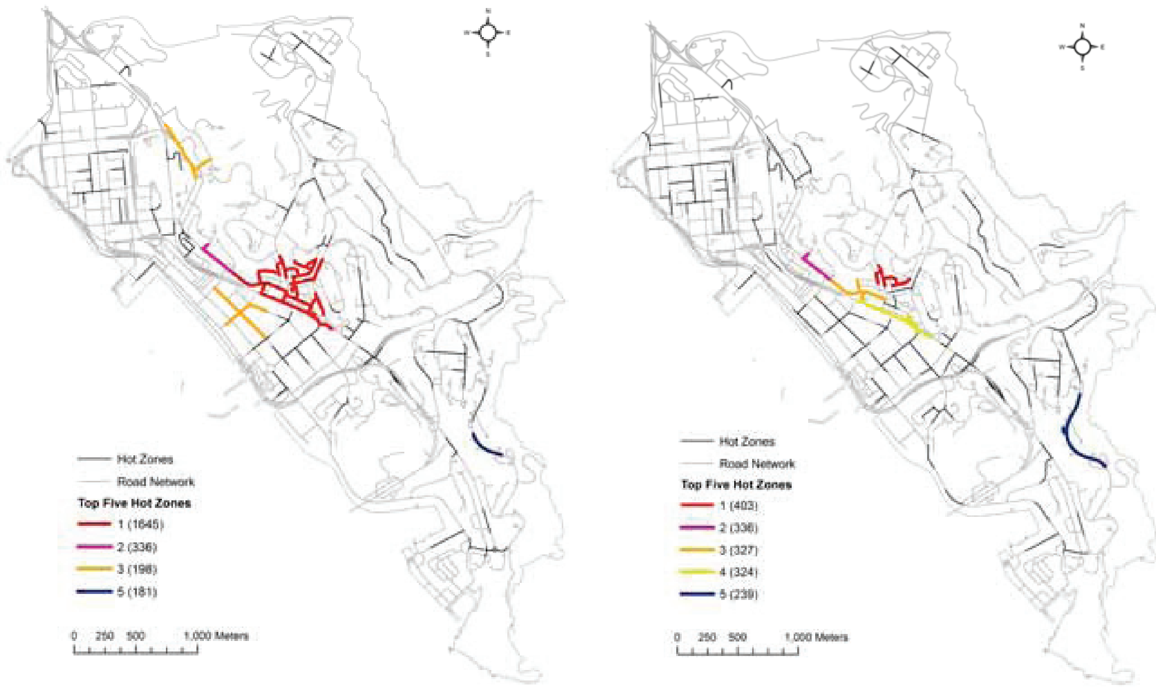
Table 4. Statistics on pedestrian hot zones

Hot Zone Type	Base model	Full model
Number of Hot Zones	50	60
Number of BSUs	221	239
Length of hot zone (km)	20.79	22.8
Number of casualties within hot zone	624	617
Costs	5067	4856

5. Conclusion

This research applies the hot zone methodology to identify dangerous road locations for pedestrians during the study period from 2005 to 2007 in Kwun Tong District of Hong Kong. In particular, the crash intensity was calculated by a casualty-weighted method, which assigns different weights to different injury severity types. Two kinds of negative binomial regression models were employed to determine the threshold values. As both the base-model and full-model pedestrian hot zones are important in addressing road safety problems, both of them should be taken seriously by administrations.

Future research can be dedicated to testing the sensitivity of hot zones on weights of injury types. In addition, more sophisticated models such as empirical bayes models can be employed to calculate the threshold values. These efforts could provide more useful information for addressing pedestrian safety problems.



(a) (b)

Fig. 3. Locations of pedestrian hot zones based on (a) the base model and (b) the full model. (Notes: Numbers in brackets are the total cost-weighted crash intensity at the hot zones.)

Acknowledgements

The authors would like to thank the Research Grant Council of Hong Kong for funding this research (HKU 749210H).

References

[1] Cheng, W., Washington, S. P., 2005. Experimental Evaluation of Hotspot Identification Methods, Accident Analysis and Prevention 37, p. 870-881.
[2] Geurts, K., Wets, G., Brijs, T., Vanhoof, K.,2004. Identification and Ranking of Black Spots - Sensitivity Analysis, Statistical Methods and Safety Data Analysis and Evaluation 1897, p. 34-42.
[3] Miranda-Moreno, L. E., Labbe, A., Fu, L. P., 2007. Bayesian Multiple Testing Procedures for Hotspot Identification, Accident Analysis and Prevention 39, p. 1192-1201.
[4] Black, W. R., Thomas, I.,1998. Accidents on Belgium's Motorway: A network Autocorrelation Analysis, Journal of Transport Geography 6, p, 23-31.
[5] Flahaut, B., Mouchart, M., San, Martin E., Thomas, I, 2003. The Local Spatial Autocorrelation and the Kernel Method for Identifying Black Zones - A Comparative Approach, Accident Analysis and Prevention 35, p. 991-1004.
[6] Loo, B. P. Y., 2009. The Identification of Hazardous Road Locations: A Comparison of the Blacksite and Hot Zone Methodologies in Hong Kong, International Journal of Sustainable Transportation 3, p. 187-202.
[7] Moons, E., Brijs, T., Wets, G., 2009. Identifying Hazardous Road Locations: Hot Spots versus Hot Zones, in *"Transactions on Computational Science Vi "* Gavrilova, M. L. and Tan, C. J. K. (Eds.). Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg, p. 288-300.
[8] Loo, B. P. Y., Yao. S., 2012. Geographic Information Systems, in *"Injury Research: Theories, Methods, and Approaches"* Li, G. and Baker, S. (Eds.). Springer, New York, p. 447-463.
[9] Loo, B. P. Y., Yao, S., Wu, J., Yu, B. and Zhong, H., 2011. Identification Method of Road Hot Zone based on GIS, Journal of Traffic and Transportation Engineering (in Chinese) 11, p. 97-103.
[10]Loo, B. P. Y., 2006. Validating Crash Locations for Quantitative Spatial Analysis: A GIS-based Approach, Accident Analysis and Prevention 38, p. 879-886.
[11]Elvik, R., 2008. A Survey of Operational Definitions of Hazardous Road Locations in Some European Countries, Accident Analysis and Prevention 40, p. 1830-1835.
[12]Elvik, R., 2006. New Approach to Accident Analysis for Hazardous Road Locations, Safety Data, Analysis, and Evaluation 1953, p. 50-55.
[13]Highway Safety Plan, Center for Advanced Public Safety, State of Alabama, 2010.
[14]Mueller, B. A., Rivara, F. P., Bergman, A. B., 1988. Urban-Rural Location and the Risk of Dying in a Pedestrian-Vehicle Collision, The Journal of Trauma 28, p. 91-94.
[15]De Blaeij, A., Florax, R. J. G. M., Rietveld, P., Verhoef, E., 2003. The Value of Statistical Life in Road Safety: A Meta-Analysis, Accident Analysis

and Prevention 35, p. 973-986.
[16]Miller, T. R., 2000. Variations between Countries in Values of Statistical Life, Journal of Transport Economics and Policy 34, p. 169-188.
[17]The True Cost of Road Crashes: Valuing Life and the Cost of a Serious Injury, International Road Assessment Programme, 2007.
[18]Graham, D. J., Glaister, S., 2003. Spatial Variation in Road Pedestrian Casualties: The Role of Urban Scale, Density and Land-Use Mix, Urban Studies 40, p. 1591-1607.
[19]Levine, N., Kim, K. E., Nitz, L. H., 1995. Spatial Analysis of Honolulu Motor Vehicle Crashes: II. Zonal Generators, Accident Analysis and Prevention 27, p. 675-685.
[20]Persaud, B., Lord, D., Palmisano, J., 2002. Calibration and Tansferability of Accident Prediction Models for Urban Intersections, Transport Research Record 1784, p. 57–64.
[21]Loo, B. P. Y., Yao, S., 2010. “The impact of area deprivation on traffic casualties in Hong Kong,” 15th HKSTS International Conference. Hong Kong.